The BD2K Training Coordinating Center’s Data Science Education Platform

José Luis Ambite, Ph.D.1, Lily Fierro, S.B.1, Florian Geigl, Ph.D.1, Jonathan Gordon, Ph.D.1, Gully A.P.C. Burns, Ph.D.1, Kristina Lerman, Ph.D.1, Jeana Kamdar, MPH2, Crystal Stewart, M.A.2, Avnish Bhattrai, M.S., Xiaoyu Lei, M.S.2, Sumiko Abe, Ph.D.2, John Darrell Van Horn, M.Eng, Ph.D.2

1Information Sciences Institute, Univ. of Southern California, Marina del Rey, CA, USA; 2USC Mark and Mary Stevens Neuroimaging and Informatics Inst., Los Angeles, CA, USA

The National Institutes of Health (NIH) has made an unprecedented investment in biomedical data science through its Big Data to Knowledge (BD2K) initiative (datascience.nih.gov). Through a series of K01 career development awards, T32/T15 institutional training awards, R25 data science training research awards, as well as the training components of fifteen U54 BD2K Centers of Excellence, the NIH has placed a premium on the development of a new generation of biomedical data science professionals1. These efforts are systematically producing unique training materials which seek to introduce everything from the basics of data science to in-depth examinations of applied biomedical analytics in the investigation of health as well as disease. In addition to these BD2K training resources, a large number of high-quality data science educational resources are available online, such as massive open online courses and video lectures. Consequently, the BD2K Training Coordinating Center (TCC) is developing a data science educational resource discovery index (ERuDIte) and an online education platform with the capability to automatically discover, describe, organize, and personalize training materials for the varied learning objectives of its users.

In this system demonstration, we will first provide a tour of the core functionality of the BD2K TCC web portal (http://www.bigdatau.org), including the exploration, search, and recommendation of learning resources tailored to a user’s learning goal. Figures 1–3 show three key site features: the Concept Map, the Faceted Search Interface, and the Educational Plan Interface. The system is already available for beta testing. Second, we will describe how we use data science techniques to build the ERuDIte index that powers the web portal. This system demonstration aims to inform the audience on the educational value of http://www.bigdatau.org for the medical informatics community, while also showing the generality and extensibility of the education platform to other domains.

Figure 1: Concept Map

Figure 2: Faceted Search

Figure 3: Customizable Educational Plan to collect resources significant to individual learning goals

References
